NASA CR-183473

Copy #2

CSC/TM-78/6056

ALGORITHMS FOR DETERMINING OPTIMAL ATTITUDE SOLUTIONS

Prepared For

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
Goddard Space Flight Center
Greenbelt, Maryland

CONTRACT NAS 5-24300
Task Assignment 824

APRIL 1978

(NASA-CR-183473) ALGORITHMS FOR DETERMINING OPTIMAL ATTITUDE SOLUTIONS (Computer Sciences Corp.) 77 p

N91-70694

Unclas
00/61 0314746

CSC

COMPUTER SCIENCES CORPORATION

CSC/TM-78/6056

ALGORITHMS FOR DETERMINING OPTIMAL ATTITUDE SOLUTIONS

Prepared for

GODDARD SPACE FLIGHT CENTER

By

COMPUTER SCIENCES CORPORATION

Under

Contract NAS 5-24300
Task Assignment 824

Prepared by:

M. D. Shuster 4/17/78
Dr. M. D. Shuster Date

Approved by:

James R. Wertz 4/17/78
Dr. J. R. Wertz Date
Section Manager

F. L. Markley 4/17/78
Dr. F. L. Markley Date
Quality Assurance Reviewer

M. Joseph 4/18/78
M. Joseph Date
Department Manager

COMPUTER SCIENCES CORPORATION
SYSTEM SCIENCES DIVISION (301) 589-1545
8728 COLESVILLE ROAD · SILVER SPRING, MARYLAND 20910

April 24, 1978

National Aeronautics and Space Administration
Goddard Space Flight Center
Greenbelt, Maryland 20771

Attention: Mr. R. Nankervis
Code 581.2
Bldg. 23, Room E-423

Subject: Contract NAS 5-24300
Task Assignment 824
Magsat Fine Aspect Documentation

Dear Mr. Nankervis:

Submitted herewith is document CSC/TM-78/6056 entitled, "Algorithms for Determining Optimal Attitude Solutions", the work for which was performed under Task 824.

This document describes the attitude computation algorithm which will be incorporated in the MSAD/Magsat Fine Aspect Attitude Determination System. The algorithms in this document are applicable to all spacecraft which compute attitude from vectors.

Yours truly,

COMPUTER SCIENCES CORPORATION

M. Plett

Dr. M. E. Plett
Technical Area Manager
Attitude Systems Operation

MEP:rtg

Attachment

Mr. R. Nankervis

Distribution

cc: GSFC
R. Werking
J. Clauson

CSC
M. Joseph
J. Wertz
D. Sood
G. Lerner
R. Beard
T. Brown
M. Shuster
F. VanLandingham
J. Guffey
M. Levitas
R. Collier
Y. Hoh
M. Baker
G. Tandon
L. Markley
R. Headrick
D. Gottlieb
T. Smelkinson
Task Library

ABSTRACT

This document presents a number of algorithms for computing the optimal rotation which carries a set of reference vectors into a set of observation vectors. These algorithms, which are based on the q-method of Davenport, provide very fast means for computing accurately the optimal attitude. They are thus suited to processing data from attitude sensors which provide the direction of a known body axis (horizon scanners, Sun sensors, star cameras) but not to sensors which provide only angle data (such as a single-axis gyro). One of these algorithms, which provides very accurate results with little computation, is tested for the sensor configuration of the Magsat mission. Numerical results are presented. A possible enhancement of this algorithm is discussed.

ACKNOWLEDGMENTS

The author is grateful to P. Davenport for interesting discussions and helpful criticisms and to F. L. Markley, whose contributions are cited in the text.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

Figure

SECTION 1 - INTRODUCTION

In this document several algorithms are presented for computing the optimal rotation which carries a set of n reference vectors into a set of n corresponding observation vectors. The rotation is optimal in the sense that it minimizes a least-square loss function.

The increasingly stringent accuracy requirements in determining spacecraft attitude require the development of optimal algorithms for attitude determination. When individual sensor accuracies exceed the accuracy requirements of a mission it is possible to use simple deterministic algorithms, which have the advantage of being easy to implement and are fast computationally. A deterministic algorithm uses only the smallest amount of data necessary to determine an attitude solution. In so doing a subset of the data is discarded.

In some missions--for example Magsat, which inspired the present work--the accuracy requirement is comparable to the accuracies of the individual sensors (in this case two star cameras and a Sun sensor). One must, therefore, take all data into account to obtain the most accurate possible determination of the spacecraft attitude. In the case of Magsat each of the three sensors provides an azimuth and coelevation in the body coordinate system (the Sun or some other star) providing in all six pieces of data all of which must be used to obtain the best possible determination of the three independent parameters specifying the spacecraft attitude.

Thus, given the n observation vectors $\hat{W}_1, \ldots, \hat{W}_n$ (for Magsat $n = 3$), corresponding to the directions, as seen in the spacecraft coordinate system, of n celestial bodies, and the n reference vectors $\hat{V}_1, \ldots, \hat{V}_n$, the known directions of these bodies in the spacecraft-centered inertial coordinate system, it is possible to write for every rotation matrix R the n equations

$$\hat{W}_i = R\,\hat{V}_i + \delta\vec{W}_i \qquad i = 1, \ldots, n \tag{1-1}$$

which define the n vectors $\delta\vec{W}_i$. The problem is to find that optimal rotation R_{opt} which minimizes the magnitudes of the $\delta\vec{W}_i$. This is the central problem of this document.

The general problem of minimizing the loss function

$$\mathcal{L}(R) = \frac{1}{2}\sum_{i=1}^{n} a_i \|\delta\vec{W}_i\|^2 \tag{1-2}$$

where the $\delta\vec{W}_i$ are given by Equation (1-1) and the a_i are positive weights, was solved by Davenport (Reference 1). In terms of the rotation matrix the minimization of $\mathcal{L}(R)$ is a formidable problem since R contains nine unknown quantities subject to six constraints. Davenport removed this difficulty by expressing R in terms of the related quaternion $\bar{q}$, which has only four components subject to the single constraint that the sum of the squares of the components be unity. Since R is a bilinear form in $\bar{q}$, $\mathcal{L}(R)$ also becomes a bilinear form in $\bar{q}$. The minimization of a bilinear form subject to a bilinear constraint is a classical problem in modern mathematics which was first solved by Euler. The solution satisfies an eigenvalue problem. The development of this eigenvalue problem leads to Davenport's q-algorithm (Section 2), to further developments of which this document is devoted.

Optimal algorithms have the obvious advantage of yielding more accurate results than deterministic algorithms. Their disadvantage is that they are sometimes computationally much slower.

Simplifications of Davenport's q-method which are fast computationally without sacrificing accuracy, are possible because in practice an optimal rotation can be chosen to make the $\delta\vec{W}_i$ very small, i.e.,

$$\|\delta\vec{W}_i(R_{opt})\| << 1 \qquad i = 1, \ldots, n \tag{1-3}$$

This is not a result of Davenport's q-method but a statement of the nature of the input reference and observation vectors. It is to be expected that $R\hat{V}_i$ can be made to overlap with $\hat{W}_i$ to within the accuracy of the sensors. For Magsat this accuracy is typically 10 arc-seconds or in natural units 5×10^{-5} radians. This is certainly very small and the techniques introduced in the present work amount to a Taylor expansion in this small quantity.

Section 2 of this document presents Davenport's derivation of the q-algorithm as first presented by Keat (Reference 2).

Section 3 of this document derives a general expression for the optimal rotation (or rather, quaternion) when the angle of rotation is infinitesimally small, i.e., when terms proportional to the square or higher powers of this angle can be neglected. For the case where there are only two observation vectors, a still simpler formula can be obtained.

In Section 4 the results of Section 3 are generalized to arbitrary rotations. A special formula is derived for the optimal quaternion which is most amenable to automatic computation.

Section 5 presents more accurate methods for determining the optimal quaternion which are necessary when the angle of rotation is very close to 180 degrees. The nature of the exact solutions to Davenport's eigenvalue problem are discussed. A Rayleigh-Schroedinger Perturbation expansion for Davenport's overlap eigenvalue is developed. Several iterative methods are also presented.

The suitability of these algorithms in different situations is discussed in Section 6. The most accurate and efficient algorithm for the purposes of definitive attitude determination in the Magsat mission is selected. Numerical results are presented for cases corresponding to the Magsat sensor configuration.

A summary of the analysis for the algorithm proposed for Magsat is presented in Section 7.

SECTION 2 - DAVENPORT'S q-ALGORITHM

Davenport (Reference 1) has developed an algorithm for determining optimum solutions for the attitude given n observation vectors and n reference vectors. This algorithm has been examined by J. Keat (Reference 2), who analyzed the algorithm for use in High Energy Astronomical Observatory-1 (HEAO-1) mission software. The derivation of Davenport's algorithm which is presented here follows closely the methods and notation of Keat's document.

2.1 STATEMENT OF THE PROBLEM

Let $\hat{W}_1, \hat{W}_2, \ldots, \hat{W}_n$ be a set of n observation unit vectors and $\hat{V}_1, \hat{V}_2, \ldots, \hat{V}_n$ be a set of n reference unit vectors. That these are unit vectors is denoted by the caret over each symbol. Let $a_1, a_2, \ldots, a_n$ be a set of n (positive) weights. For any rotation matrix R, the loss function $\ell(R)$ is defined as

$$\ell(R) = \frac{1}{2} \sum_{i=1}^{n} a_i \left\| \hat{W}_i - R\hat{V}_i \right\|^2 \qquad (2\text{-}1)$$

The rotation matrix R_{opt} which minimizes the loss function is said to be the optimal rotation (in a least-squares sense) which carries the vectors $\hat{V}_i$, $i = 1, \ldots, n$, into the vectors $\hat{W}_i$, $i = 1, \ldots, n$.

In actual practice $\hat{V}_i$, $i = 1, \ldots, n$, may be the position of n stars in some reference frame (for example, the spacecraft-centered inertial frame) and $\hat{W}_i$, $i = 1, \ldots, n$, the observed position of these stars in a body-fixed coordinate system. The a_i are the relative weights of the observations determined by the quality of the measuring apparatus. R_{opt} is then the optimal (least-squares) estimate of the rotation which carries the inertial axes into the body axes and, thus, provides an optimal estimate of the attitude of the body.

Since the loss function may be scaled without affecting the result for the optimal estimate of the rotation, it is possible to set

$$\sum_{i=1}^{n} a_i = 1 \tag{2-2}$$

which will be assumed to be true throughout this work.

It follows that

$$\ell(R) = 1 - g(R) \tag{2-3}$$

where

$$g(R) = \sum_{i=1}^{n} a_i \hat{W}_i \cdot R \hat{V}_i \tag{2-4}$$

is called the gain function. $\ell(R)$ is a minimum if and only if $g(R)$ is a maximum. Henceforth, all attention will be directed toward finding R which maximizes $g(R)$.

2.2 MATRIX REPRESENTATION OF THE GAIN FUNCTION

It is convenient to define new vectors $\vec{V}_i$, $i = 1, \ldots, n$, and $\vec{W}_i$, $i = 1, \ldots, n$, according to

$$\vec{W}_i = \sqrt{a_i}\, \hat{W}_i \qquad i=1, \ldots, n \tag{2-5a}$$

$$\vec{V}_i = \sqrt{a_i}\, \hat{V}_i \qquad i=1, \ldots, n \tag{2-5b}$$

whence

$$g(R) = \sum_{i=1}^{n} \vec{W}_i \cdot R\vec{V}_i \tag{2-6}$$

Defining $3 \times n$ matrices according to

$$W = [\vec{W}_1, \vec{W}_2, \ldots, \vec{W}_n] \tag{2-7a}$$

$$V = [\vec{V}_1, \vec{V}_2, \ldots, \vec{V}_n] \tag{2-7b}$$

the $n \times n$ matrix $W^T RV$ may be formed. Here W^T, an $n \times 3$ matrix, is the transpose of W. The (i, j) component of this matrix is just $\vec{W}_i \cdot R\vec{V}_j$. Thus,

$$g(R) = T_r[W^T R V] \tag{2-8}$$

where Tr denotes the trace operation.

Note that $\vec{V}_i$ may be interpreted as a 3×1 matrix, whose transpose, $\vec{V}_i^T$, is a 1×3 matrix. In this notation

$$\vec{W}_i \cdot R\vec{V}_j = \vec{W}_i^T R\vec{V}_j \tag{2-9}$$

and the dyadic (Kronecker product) is written $\vec{V}_i\vec{W}_j^T$. The two notations in Equation (2-9) will be used interchangeably.

Since the trace of a product of matrices is unchanged by a cyclic reordering of the factors it follows that

$$g(R) = T_r \left[R V W^T \right] \tag{2-10}$$

or,

$$g(R) = T_r \left[R B^T \right] \tag{2-11}$$

where

$$B = W V^T \tag{2-12}$$

$$= \sum_{i=1}^{n} \vec{W}_i \vec{V}_i^{\,T} \tag{2-13}$$

$$= \sum_{i=1}^{n} a_i \hat{W}_i \hat{V}_i^{\,T} \tag{2-14}$$

2.3 THE GAIN FUNCTION IN QUATERNION FORM

The quaternion $\bar{q}$ representing a rotation is a four-dimensional vector given by

$$\bar{q}_{4\times 1} = \begin{Bmatrix} \vec{Q}_{3\times 1} \\ q_{1\times 1} \end{Bmatrix} = \begin{Bmatrix} \hat{X} \sin(\theta/2) \\ \cos(\theta/2) \end{Bmatrix} \tag{2-15}$$

where $\hat{X}$ is the axis of rotation and θ is the angle of rotation about $\hat{X}$. Note

$$\|\bar{q}\|^2 = \bar{q}^T \bar{q} = \sum_{i=1}^{4} q_i^2 = |\vec{Q}|^2 + q^2 = 1 \tag{2-16}$$

The relationship between R and $\bar{q}$ (sometimes called the Euler symmetric parameters in this form) is well known (Reference 3).

$$R(\bar{q}) = (q^2 - \vec{Q}\cdot\vec{Q})I + 2\vec{Q}\vec{Q}^T - 2q\underset{\sim}{\vec{Q}} \tag{2-17}$$

where I is the identity matrix and $\underset{\sim}{\vec{Q}}$ is a skew-symmetric matrix given by

$$[\underset{\sim}{\vec{Q}}] = \begin{bmatrix} 0 & -Q_3 & Q_2 \\ Q_3 & 0 & -Q_1 \\ -Q_2 & Q_1 & 0 \end{bmatrix} \tag{2-18}$$

or, in terms of the Levi-Civita symbol, ϵ_{ijk}, defined to be completely anti-symmetric in the indices i, j, k with $\epsilon_{123} = 1$)

$$[\underset{\sim}{\vec{Q}}]_{ij} = -\sum_k \epsilon_{ijk} Q_k \tag{2-19}$$

In terms of the quaternion $\bar{q}$ the gain function may be written as

$$\begin{aligned} g(\bar{q}) &= g(R(\bar{q})) \\ &= (q^2 - \vec{Q}\cdot\vec{Q})\,\mathrm{Tr}\,B^T + 2\,\mathrm{Tr}(\vec{Q}\vec{Q}^T B^T) - 2q\,\mathrm{Tr}[\underset{\sim}{\vec{Q}} B^T] \end{aligned} \tag{2-20}$$

It is convenient to define scalar, matrix, and vector quantities σ , S , $\vec{Z}$ according to

$$\sigma = \mathrm{Tr}\, B = \sum_{i=1}^{n} a_i (\hat{W}_i \cdot \hat{V}_i) \tag{2-21}$$

$$S = B + B^T = \sum_{i=1}^{n} a_i [\hat{W}_i \hat{V}_i^T + \hat{V}_i \hat{W}_i^T] \tag{2-22}$$

$$\vec{Z} = \sum_{i=1}^{n} a_i (\hat{W}_i \times \hat{V}_i) \tag{2-23}$$

Equation (2-23) is equivalent to

$$\underset{\sim}{\vec{Z}} = B^T - B \tag{2-24}$$

In terms of these quantities

$$\begin{aligned} 2\, \mathrm{Tr}\, (\vec{Q} \vec{Q}^T B^T) &= 2\, \mathrm{Tr}\, (\vec{Q}^T B^T \vec{Q}) \\ &= 2\, \mathrm{Tr}\, (\vec{Q}^T B \vec{Q}) \\ &= \mathrm{Tr}\, (\vec{Q}^T (B + B^T) \vec{Q}) \\ &= \vec{Q}^T S \vec{Q} \end{aligned} \tag{2-25}$$

$$\begin{aligned} \mathrm{Tr}\, (\underset{\sim}{\vec{Q}} B^T) &= \sum_{jk} [\underset{\sim}{\vec{Q}}]_{jk} B_{jk} \\ &= \sum_{jk} \sum_{\ell} \left\{ -\epsilon_{\ell jk} Q_\ell \sum_i a_i (\hat{W}_i)_j (\hat{V}_i)_k \right\} \\ &= -\sum_{\ell} Q_\ell \sum_i a_i \epsilon_{\ell jk} (\hat{W}_i)_j (\hat{V}_i)_k \end{aligned} \tag{2-26}$$

$$= -\sum_{\ell} Q_{\ell} \sum_{i} a_i \, (\hat{w}_i \times \hat{v}_i)_{\ell}$$

$$= \; -\vec{Q} \cdot \vec{Z} \qquad \text{(2-26) (Cont'd)}$$

Using these results

$$g(\bar{q}) = \sigma(q^2 - \vec{Q} \cdot \vec{Q}) + \vec{Q}^T S \vec{Q} + 2q\,\vec{Q} \cdot \vec{Z} \qquad \text{(2-27)}$$

A simple rearrangement yields

$$g(\bar{q}) = \bar{q}^T K \bar{q} \qquad \text{(2-28)}$$

where

$$K = \left[\begin{array}{c|c} S - \sigma I & \vec{Z} \\ \hline \vec{Z}^T & \sigma \end{array}\right] \qquad \text{(2-29)}$$

2.4 DETERMINATION OF THE LEAST-SQUARES ATTITUDE SOLUTION

The problem of obtaining the optimal estimate of the attitude has been reduced to finding the quaternion $\bar{q}$ which maximizes the gain function of Equation (2-28) subject to the constraint

$$\bar{q}^T \bar{q} = 1 \qquad \text{(2-30)}$$

The constraint is taken into account most easily by the method of Lagrange multipliers. A new gain function $g'(\bar{q})$ is defined

$$g'(\bar{q}) = \bar{q}^T K \bar{q} - \lambda \bar{q}^T \bar{q} \qquad \text{(2-31)}$$

$g'(\bar{q})$ is now maximized without constraint and the unknown Lagrange multiplier λ is determined such that Equation (2-30) is satisfied. Straightforward differentiation establishes that $g'(\bar{q})$ attains a stationary point provided

$$K\bar{q} = \lambda \bar{q} \tag{2-32}$$

Thus, λ is an eigenvalue of K. Since K is real symmetric $(K^T = K)$ λ is always real. Examination of Equation (2-28) shows that λ must also be the largest of the four eigenvalues of K since the maximum value of $g(\bar{q})$ is sought. The constraint, Equation (2-30), can always be satisfied since Equation (2-32) does not determine the norm of $\bar{q}$. This is Davenport's algorithm for determining the optimal least-squares attitude.

It may be noted in passing that since $\mathrm{Tr}\, K = 0$ the sum of the eigenvalues of K must also be zero.

SECTION 3 - OPTIMAL INFINITESIMAL ROTATIONS

A simple and elegant solution obtains in the case where the observation vectors differ from their respective reference vectors by a very small amount. In this case the optimal rotation is infinitesimal, i.e., the angle of the rotation carrying the reference vectors into the observation vectors is so small that terms proportional to the square or higher powers of that angle can be neglected within the accuracy requirements of the attitude solution. In this section the general form of the solution where the angle of rotation is infinitesimal (in the above sense) is derived. It will be shown that the greatest eigenvalue of K in the infinitesimal case differs from unity by only very small terms. This allows the four dimensional eigenvalue problem to be converted approximately into a system of three linear equations.

3.1 THE INFINITESIMAL K-MATRIX

Since the angle of rotation is assumed to be infinitesimal, it is possible to write

$$\hat{W}_i = (I + \varepsilon_i \Omega_i)\hat{V}_i + O(\varepsilon_i^2), \quad i = 1, \ldots, n \tag{3-1}$$

with Ω_i a skew-symmetric matrix which is O(1) and ϵ_i is very small compared to unity. In the notation of Equation (2-18)

$$\Omega_i = -\underset{\sim}{\hat{X}}_i \tag{3-2}$$

where $\hat{X}_i$ is the ith axis of rotation (a unit vector) and

$$|\varepsilon_i| << 1 \tag{3-3}$$

By $O(\epsilon^k)$ is understood a quantity which tends to zero as a constant $\times\, \epsilon^k$ as $\epsilon \to 0$. In the following

$$\varepsilon = \max \{ |\varepsilon_i| \,;\, i = 1, \ldots, n \} \tag{3-4}$$

It should be noted that the $\hat{W}_i$ and $\hat{V}_i$ do not determine ϵ_i and Ω_i (or equivalently ϵ_i and $\hat{X}_i$) uniquely. The assumption, however, that $\hat{W}_i$ and $\hat{V}_i$ differ by an infinitesimal amount implies that it is possible to choose Ω_i (or $\hat{X}_i$) such that $|\epsilon_i| << 1$. Clearly, ϵ_i will achieve the smallest possible magnitude when $\hat{X}_i$ is chosen parallel to $\hat{W}_i \times \hat{V}_i$.

In the infinitesimal case

$$S = 2 \sum_{i=1}^{n} a_i \hat{W}_i \hat{W}_i^T + \sum_i \varepsilon_i a_i [\Omega_i, \hat{W}_i \hat{W}_i^T] + O(\varepsilon^2) \tag{3-5}$$

where $[A, B] = AB - BA$ denotes the commutator of A and B. Since the trace of a commutator vanishes

$$\begin{aligned} \sigma = \mathrm{Tr}\, \mathcal{B} &= \tfrac{1}{2} \mathrm{Tr}\, S \\ &= \sum_{i=1}^{n} a_i + O(\varepsilon^2) \\ &= 1 + O(\varepsilon^2) \end{aligned} \tag{3-6}$$

Note also

$$\vec{Z} = \sum_{i=1}^{n} a_i \hat{W}_i \times \hat{V}_i = O(\varepsilon) \tag{3-7}$$

The fact that $\sigma = 1 + O(\epsilon^2)$ and $\vec{Z} = O(\epsilon)$ will lead to the result that $\lambda_{max} = 1 + O(\epsilon^2)$ where λ_{max} is the largest eigenvalue of K . This will be proved in two steps.

3.2 A LEMMA FOR THE NULL ROTATION

The following lemma will be proved:

Lemma: For the null rotation ($\epsilon = 0$) $\lambda_{max} = 1$ and is a simple eigenvalue (non-degenerate) provided that the reference vectors $\hat{V}_i$ $i = 1, \ldots, n$, are not all collinear.

Proof: $\epsilon = 0$ implies that $\epsilon_i = 0$, $i = 1, \ldots, n$. In this case

$$\sigma = 1 \tag{3-8}$$

$$S = S_o = 2\sum_{i=1}^{n} a_i \hat{V}_i \hat{V}_i^T \tag{3-9}$$

$$\vec{Z} = \vec{0} \tag{3-10}$$

K then has the form

$$K = K_o = \left[\begin{array}{c|c} S_o - I & \vec{0} \\ \hline \vec{0}^T & 1 \end{array}\right] \tag{3-11}$$

so that 1 is certainly an eigenvalue of K . To show that it is the largest eigenvalue and non-degenerate the eigenvalues of $S_o - I$ must be determined. Examine the eigenvalues of S_o . Let $\hat{v}$ be a normalized eigenvector of S_o with eigenvalue μ . Then

$$S_o \hat{v} = \mu \hat{v} \tag{3-12}$$

and

$$\mu = \hat{v}^T S_0 \hat{v} \tag{3-13}$$

$$= 2\sum_{i=1}^{n} a_i (\hat{V}_i \cdot \hat{v})^2 \tag{3-14}$$

The expression on the right is non-negative; hence, $\mu \geq 0$. Likewise,

$$2\sum_{i=1}^{n} a_i (\hat{V}_i \cdot \hat{v})^2 \leq 2\sum_{i=1}^{n} a_i = 2 \tag{3-15}$$

so that

$$0 \leq \mu \leq 2 \tag{3-16}$$

μ attains the value 2 if and only if

$$\hat{V}_i = \pm \hat{v} \qquad i = 1, \ldots, n \tag{3-17}$$

Likewise the eigenvalues λ' of $S_o - I$ satisfy

$$-1 \leq \lambda' \leq +1 \tag{3-18}$$

and $\lambda' = +1$ if and only if Equation (3-17) is satisfied. This proves the lemma.

It may be noted that the λ' will be significantly less than unity for the null rotation provided that the $\hat{V}_i$ are significantly different from one another. This point will be examined in greater detail for the case of two observations at the end of this section.

3.3 DETERMINATION OF THE OPTIMAL INFINITESIMAL ROTATION

Examine now the characteristic polynomial for K when $\epsilon \neq 0$. Setting

$$\vec{Z} = \varepsilon \vec{z} \tag{3-19}$$

with $|\vec{z}| = O(1)$, K can be written as

$$K = \left[\begin{array}{c|c} S - I & \varepsilon \vec{z} \\ \hline \varepsilon \vec{z}^{\,T} & 1 \end{array}\right] + O(\varepsilon^2) \tag{3-20}$$

Without loss of generality the coordinate system may be chosen so that $\vec{z} = |\vec{z}|\, \hat{x}$. Then the determinant defining the characteristic equation for λ may be expanded in minors to give

$$0 = \det |K - \lambda I|$$

$$= (1-\lambda) \det |S - (\lambda+1) I| - \varepsilon |\vec{z}| \det \begin{vmatrix} S_{12} & S_{13} & \varepsilon|\vec{z}| \\ S_{22}-1-\lambda & S_{23} & 0 \\ S_{32} & S_{33}-1-\lambda & 0 \end{vmatrix} + O(\varepsilon^2) \tag{3-21}$$

$$= (1-\lambda) \det |S - (\lambda+1) I| - \varepsilon^2 |\vec{z}|^2 \det \begin{vmatrix} S_{22}-1-\lambda & S_{23} \\ S_{32} & S_{33}-1-\lambda \end{vmatrix} + O(\varepsilon^2)$$

Thus, it follows that λ is a solution of

$$(1-\lambda) f_0(\lambda) + \varepsilon^2 f_1(\lambda) = 0 \tag{3-22}$$

from which

$$\lambda_{max} = 1 + O(\varepsilon^2) \tag{3-23}$$

The same result will not hold for the three smaller eigenvalues of K , which will, in general, contain corrections to the "unperturbed" values of order ϵ.

Examine now

$$K \bar{\xi} = \lambda_{max} \bar{\xi} \tag{3-24}$$

The first three components of this equation are

$$(S - \sigma I) \vec{Q} + \vec{Z} \xi = \lambda_{max} \vec{Q} \tag{3-25}$$

which leads to

$$\vec{Q} = -(S - (\sigma + \lambda_{max}) I)^{-1} \vec{Z} \xi \tag{3-26}$$

Recalling

$$\sigma = 1 + O(\varepsilon^2)$$

$$\lambda_{max} = 1 + O(\varepsilon^2)$$

$$\xi = 1 + O(\varepsilon^2)$$

$$\vec{Z} = O(\varepsilon)$$

it follows that

$$\vec{Q} = -(S-2I)^{-1}\vec{Z} + O(\varepsilon^2) \tag{3-27}$$

From $S = S_o + O(\epsilon)$ a slightly poorer approximation also results:

$$\vec{Q} = -(S_o-2I)^{-1}\vec{Z} + O(\varepsilon^2) \tag{3-28}$$

These are the desired algorithms.

Implicit in the above equations is the assumption that $S_o - 2I$ and $S - 2I$ are non-singular. Since these two matrices differ by an amount of order ϵ, it is sufficient to show that $S_o - 2I$ has no eigenvalue of infinitesimal size.

Denote the three smaller eigenvalues of K_o by $\lambda_2^{(o)}$, $\lambda_3^{(o)}$, $\lambda_4^{(o)}$. Then the eigenvalues of $S_o - 2I$ are $\lambda_i^{(o)} - 1$, $i = 2, 3, 4$. For the case of only two observations the three eigenvalues of $S_o - 2I$ are given by

$$\lambda_2^{(o)} - 1 = +\sqrt{1 - 4a_1a_2|\hat{V}_1 \times \hat{V}_2|^2} - 1 \tag{3-29a}$$

$$\lambda_3^{(o)} - 1 = -\sqrt{1 - 4a_1a_2|\hat{V}_1 \times \hat{V}_2|^2} - 1 \tag{3-29b}$$

$$\lambda_4^{(o)} - 1 = -2 \tag{3-29c}$$

none of which vanish.

For the special case of equal weights $(a_1 = a_2 = 1/2)$ these become

$$\lambda_2^{(0)} - 1 = (\hat{V}_1 \cdot \hat{V}_2) - 1 \tag{3-30a}$$

$$\lambda_3^{(0)} - 1 = -(\hat{V}_1 \cdot \hat{V}_2) - 1 \tag{3-30b}$$

$$\lambda_4^{(0)} - 1 = -2 \tag{3-30c}$$

which are not infinitesimally small provided that $|\hat{V}_1 \cdot \hat{V}_2| = |\hat{W}_1 \cdot \hat{W}_2|$ is not infinitesimally close to unity. Thus, an important condition on the usefulness of Equations (3-27) and (3-28) is that

$$1 - |\hat{W}_1 \cdot \hat{W}_2| >> \varepsilon \tag{3-31}$$

If Equation (3-31) is not satisfied then $\vec{Q}$ as defined by Equations (3-27) or (3-28) will contain large admixtures of unwanted attitude solutions.

Expressed in terms of the determinant of $S_0 - 2I$ the condition becomes

$$|\det(S_0 - 2I)| = 8a_1a_2|\hat{W}_1 \times \hat{W}_2|^2 >> \varepsilon \tag{3-32}$$

The essential content of Equations (3-31) and (3-32) is that the solution for the attitude is most accurate when the observations are orthogonal. This is true for the exact solution of Equation (2-32) as well.

3.4 A SPECIAL FORMULA FOR THE CASE OF TWO OBSERVATIONS

Equations (3-27) and (3-28) are a great simplification of the exact Equation (2-32) replacing the eigenvalue problem for a 4×4 matrix by the inversion of a

3×3 matrix. For the case of only two observations Equation (3-28) has a still simpler form.

To order ϵ^2 the vector components of the quaternion representing an optimal infinitesimal rotation is given by

$$\vec{Q} = \frac{1}{2}\left(I - \frac{1}{2} S_0\right)^{-1} \vec{Z} \tag{3-33}$$

which follows directly from Equation (3-28). The evaluation of Equation (4-1) will be greatly simplified if a simple form can be found for the inverse of $I - \frac{1}{2} S_o$.

For n observations

$$I - \frac{1}{2} S_0 = I - \sum_{i=1}^{n} a_i \hat{W}_i \hat{W}_i^T \tag{3-34}$$

Since the algebra of $\{I;\ W_i W_j^T,\ i, j = 1, \ldots, n\}$ is closed, it is possible to write the inverse of $I - \frac{1}{2} S_o$ as

$$\left(I - \frac{1}{2} S_0\right)^{-1} = I + \sum_{ij} P_{ij} \hat{W}_i \hat{W}_j^T \tag{3-35}$$

The coefficients P_{ij} may be determined by insisting that

$$I = \left(I - \sum_{i=1}^{n} a_i \hat{W}_i \hat{W}_i^T\right)\left(I + \sum_{ij} P_{ij} \hat{W}_i \hat{W}_j^T\right) \tag{3-36}$$

Noting that

$$\hat{w}_i \hat{w}_i^T \hat{w}_j \hat{w}_k^T = (\hat{w}_i \cdot \hat{w}_j) \hat{w}_i \hat{w}_k^T \tag{3-37}$$

Equation (3-36) becomes

$$I = I + \sum_{ij} \left\{ -a_i \delta_{ij} + P_{ij} - \sum_k a_i (\hat{w}_i \cdot \hat{w}_k) P_{kj} \right\} \hat{w}_i \hat{w}_j^T \tag{3-38}$$

or

$$-a_i \delta_{ij} + P_{ij} - \sum_k a_i (\hat{w}_i \cdot \hat{w}_k) P_{kj} = 0 \tag{3-39}$$

Defining $n \times n$ matrices A and B according to

$$A_{ij} = a_i \delta_{ij} \tag{3-40a}$$

$$B_{ij} = a_i (\hat{w}_i \cdot \hat{w}_j) \tag{3-40b}$$

(B here should not be confused with B of Section 2) leads to the solution in matrix notation

$$P = (I - B)^{-1} A \tag{3-41}$$

Thus the inversion of a 3×3 matrix has been replaced by the inversion of an $n \times n$ matrix. Clearly, this is of practical importance only for the case $n = 2$. For that case

$$(I - B) = \begin{bmatrix} 1 - a_1 & -a_1(\hat{w}_1 \cdot \hat{w}_2) \\ -a_2(\hat{w}_1 \cdot \hat{w}_2) & 1 - a_2 \end{bmatrix} \tag{3-42}$$

for which the inverse is

$$(I - B)^{-1} = \frac{1}{|\hat{w}_1 \times \hat{w}_2|^2} \begin{bmatrix} a_2^{-1} & a_2^{-1}(\hat{w}_1 \cdot \hat{w}_2) \\ a_1^{-1}(\hat{w}_1 \cdot \hat{w}_2) & a_1^{-1} \end{bmatrix} \tag{3-43}$$

as may be easily verified. From Equations (3-41) and (3-43) it then follows that

$$P = \frac{1}{|\hat{w}_1 \times \hat{w}_2|^2} \begin{bmatrix} a_1/a_2 & (\hat{w}_1 \cdot \hat{w}_2) \\ (\hat{w}_1 \cdot \hat{w}_2) & a_2/a_1 \end{bmatrix} \tag{3-44}$$

Recalling

$$\vec{Q} = \tfrac{1}{2}\left(I - \sum_{ij} P_{ij}\, \hat{w}_i \hat{w}_j^T\right) \vec{Z} + O(\varepsilon^2) \tag{3-45}$$

$$\vec{Z} = \sum_i a_i (\hat{w}_i \times \hat{v}_i) \tag{3-46}$$

leads immediately to

$$
\begin{aligned}
\vec{Q} = \frac{1}{2}\Big\{ & a_1 \hat{w}_1 \times \hat{v}_1 + a_2 \hat{w}_2 \times \hat{v}_2 \\
& + a_1 \left[(a \hat{v}_2 - b \hat{v}_1) \cdot (\hat{w}_1 \times \hat{w}_2) \right] \hat{w}_1 \\
& + a_2 \left[(b \hat{v}_2 - a \hat{v}_1) \cdot (\hat{w}_1 \times \hat{w}_2) \right] \hat{w}_2 \Big\} \\
& + O(\varepsilon^2)
\end{aligned}
\tag{3-47}
$$

with a and b given by

$$
a = \frac{1}{|\hat{w}_1 \times \hat{w}_2|^2} \tag{3-48a}
$$

$$
b = \frac{(\hat{w}_1 \cdot \hat{w}_2)}{|\hat{w}_1 \times \hat{w}_2|^2} \tag{3-48b}
$$

SECTION 4 - APPROXIMATE OPTIMAL SOLUTIONS FOR FINITE ROTATIONS

The result of the last section may be generalized with very little change to apply to finite rotations as well.

Recall Equation (3-26) which read

$$\vec{Q} = -\left(S - (\sigma + \lambda_{max})I\right)^{-1} \vec{Z}\, q \tag{4-1}$$

This equation is true for an arbitrary solution to Equation (3-24) provided that the matrix $(S - (\sigma + \lambda_{max}) I)$ is nonsingular. It will be shown in Section 5.1 that this is true so long as the angle of rotation is not π.

In terms of the Gibbs' vector, $\vec{Y}$,

$$\vec{Y} = \vec{Q}/q \tag{4-2}$$

Equation (4-1) becomes

$$\vec{Y} = -\left(S - (\sigma + \lambda_{max})I\right)^{-1} \vec{Z} \tag{4-3}$$

All the quantities appearing in Equation (4-3) are immediately obtainable except λ_{max}. Thus, the specification of $\vec{Y}$ is limited only by the uncertainty in the overlap eigenvalue λ_{max}. Once $\vec{Y}$ is known, $\bar{q}$ may be determined from

$$\vec{Q} = \vec{Y}/\sqrt{1 + |\vec{Y}|^2} \tag{4-4a}$$

$$q = 1/\sqrt{1 + |\vec{Y}|^2} \tag{4-4b}$$

4.1 THE OVERLAP EIGENVALUE FOR FINITE ROTATIONS

We now prove the following key result.

Let $\hat{V}_i$, $i = 1, \ldots, n$, and $\hat{W}_i$, $i = 1, \ldots, n$, be sets of reference and observation unit vectors, respectively, and let it be supposed that it is possible to find a single rotation matrix R_m such that

$$\hat{W}_i = R_m \hat{V}_i + O(\varepsilon_i) \qquad i = 1, \ldots, n \tag{4-5}$$

with

$$|\varepsilon_i| << 1 \qquad i = 1, \ldots, n \tag{4-6}$$

Let

$$\varepsilon = \max \{ |\varepsilon_i| , i = 1, \ldots, n \} \tag{4-7}$$

Then

$$\lambda_{max} = 1 + O(\varepsilon^2) \tag{4-8}$$

whatever may be the value of the angle of rotation characterising R_m .

The proof of this result in the special case that each $\hat{W}_i$ differed from $\hat{V}_i$ by only an infinitesimal rotation was the work of Section 3. In that case R_m was the identity matrix. The result will be used again here.

That there exists a matrix R_m which minimizes the differences $|\hat{W}_i - R_m \hat{V}_i|$ in a least squares sense was shown by Davenport's algorithm (Section 2). That these differences will be infinitesimal is not a result of Davenport's algorithm but rather an assumption on the nature of the $\hat{W}_i$ and $\hat{V}_i$ as pointed out in the introduction (Section 1).

Recall that the gain function $g(\bar{q})$ given by

$$g(\bar{q}) = \bar{q}^T K \bar{q} \tag{4-9}$$

is identical in value to

$$g(R) = \sum_{i=1}^{n} a_i \hat{W}_i \cdot R \hat{V}_i \tag{4-10}$$

provided that R and $\bar{q}$ are related by Equation (2-17).

If $\bar{q}_m$ maximizes the right member of Equation (4-9) then, clearly, by Equations (2-30) and (2-32)

$$g(\bar{q}_m) = \lambda_{max} \tag{4-11}$$

and the same is true for the related equation for R_m

$$g(R_m) = \lambda_{max} \tag{4-12}$$

Consider now the case where instead of the set of reference vectors $\hat{V}_i$, $i = 1, \ldots, n$, a different set is chosen, namely

$$\hat{U}_i = R_o \hat{V}_i \qquad i = 1, \ldots, n \tag{4-13}$$

where R_o is an arbitrary rotation matrix. The observation vectors remain the same.

The gain function $g'(R)$ for the rotation matrix R which carries the set $\hat{U}_i$ into the set $\hat{V}_i$ is taken to be

$$g'(R) = \sum_{i=1}^{n} a_i \hat{W}_i \cdot R \hat{U}_i \tag{4-14}$$

Obviously, the rotation matrix R'_m which maximizes this $g'(R)$ is

$$R'_m = R_m R_o^{-1} \tag{4-15}$$

and

$$g'(R'_m) = \lambda_{max} \tag{4-16}$$

with λ_{max} having the same value as in Equations (4-11) and (4-12) as may be verified by direct substitution.

Now let

$$R_o = R_m \tag{4-17}$$

which does not affect the value of λ_{max}. For this choice of R_o

$$\hat{W}_i = \hat{U}_i + O(\varepsilon_i) \quad i = 1, \ldots, n \tag{4-18}$$

with the same ϵ_i as in Equation (4-5).

But this is just the case of the infinitesimal rotation of Section 3, so that it must follow that

$$\lambda_{max} = 1 + O(\varepsilon^2) \tag{4-19}$$

independent of the size of the angle of rotation of the optimal rotation. (For the $\hat{U}_i$, $i = 1, \ldots, n$, of Equation (4-13) with $R_o = R_m$ it follows that R'_m is the identity matrix corresponding to a vanishing angle of rotation of the optimum rotation.)

From the practical standpoint the order of the deviation of λ_{max} from unity is determined only by the order of the computational and experimental errors in the reference and observation vectors and not of the actual angle of rotation.

4.2 A SPECIAL FORMULA FOR THE COMPUTATION OF FINITE OPTIMAL ROTATIONS

Equation (3-47) can hold only for infinitesimal rotations since it replaces the matrix S by S_o. Markley (Reference 4), however, has offered a useful result for finite rotations which reduces to Equation (3-47) when the optimal rotation is infinitesimal.

The problem is to invert the matrix

$$[(\omega + \sigma)I - S]$$

for an arbitrary 3×3 matrix S. We show with Markley that the inverse of this matrix may be expressed as a quadratic polynomial in S. The form of the constant multiplying I is arbitrary and has been chosen to agree in form with Equation (4-1).

Markley's formula may be derived as follows.

If ξ is an eigenvalue of S , then the characteristic equation for ξ is

$$\begin{aligned} 0 &= \det |S - \xi I| \\ &= -\xi^3 + 2\sigma \xi^2 - \varkappa \xi + \Delta \end{aligned} \tag{4-20}$$

with

$$\sigma = \frac{1}{2} \mathrm{Tr}\, S \tag{4-21a}$$

$$\Delta = \det S \tag{4-21b}$$

$$\begin{aligned} \varkappa &= \mathrm{Tr}\,(\mathrm{adj}\, S) \\ &= S_{11} S_{22} + S_{22} S_{33} + S_{33} S_{11} \\ &\quad - S_{12} S_{21} - S_{23} S_{32} - S_{31} S_{13} \end{aligned} \tag{4-21c}$$

By the Cayley-Hamilton theorem S also satisfies Equation (4-20) so that

$$S^3 = 2\sigma S^2 - \varkappa S + \Delta I \tag{4-22}$$

It follows that any function of S may be written as a quadratic polynomial in S . Therefore, it is possible to find constants α , β , and γ such that

$$\left[(\omega + \sigma) I - S\right]^{-1} = \gamma^{-1} \left(\alpha I + \beta S + S^2\right) \tag{4-23}$$

α , β , and γ may be determined by the same method which was used to determine P_{ij} in Section 4.1. This leads to

$$\alpha(\omega) = (\omega^2 - \sigma^2 + \varkappa) \tag{4-24a}$$

$$\beta(\omega) = \omega - \sigma \tag{4-24b}$$

$$\gamma(\omega) = (\omega + \sigma)(\omega^2 - \sigma^2 + \varkappa) - \Delta \tag{4-24c}$$

and

$$\vec{Y} = \gamma^{-1}(\lambda_{max})[\alpha(\lambda_{max})I + \beta(\lambda_{max})S + S^2]\vec{Z} \tag{4-25}$$

Equation (4-25) is exact. Making the approximation $\lambda_{max} = 1$ leads to

$$\vec{Y} = \gamma^{-1}(1)[\alpha(1)I + \beta(1)S + S^2]\vec{Z} + O(\varepsilon^2) \tag{4-26}$$

which is the analogue of Equation (3-47) for a finite rotation and is true for any number of observation vectors provided that at least two of the weights do not vanish. It is simple, though tedious, to verify that Equation (4-26) reduces to Equation (3-47) for the case of an infinitesimal rotation. The generalization of Equation (3-47) for n observation vectors but for infinitesimal rotations is obtained by setting $\lambda_{max} \approx \sigma$ and $S \approx S_o$. Then

$$\vec{Q} = \frac{1}{2\varkappa_o - \Delta_o}[\varkappa_o I + S_o^2]\vec{Z} + O(\delta^2) \tag{4-27}$$

where κ_o and Δ_o are computed from S_o. δ is defined as the larger of ϵ and the angle of rotation. This distinction between δ and ϵ will be used henceforth to distinguish approximations which hold only for very small angles from those which hold everywhere.

A somewhat better approximation than Equation (4-27) can be had by keeping $\lambda_{max} \approx \sigma$ but not approximating S_o. This leads to

$$\vec{Q} = \frac{1}{2\sigma\kappa - \Delta}\left[\kappa I + S^2\right]\vec{Z} + O(\delta^2) \tag{4-28}$$

The small computational advantage which these last two equations have over Equation (4-26) is more than outweighed by the disadvantage of their restricted region of applicability. The approximation $\lambda_{max} \approx 1$ will never be a poorer approximation than $\lambda_{max} \approx \sigma$.

SECTION 5 - HIGHER APPROXIMATIONS FOR FINITE OPTIMAL ROTATIONS

It was remarked in the last section that the accuracy of determining an optimal rotation, whether the representation be the rotation matrix, the quaternion, or the Gibb's vector, is limited only by the accuracy with which λ_{max} can be determined. It is the purpose of this section to present more accurate expressions for λ_{max}. For most cases this is unnecessary since ϵ is already very small and ϵ^2 is clearly negligible. The difficulty is that the matrix

$$[(\lambda_{max} + \sigma) I - S]$$

is ill-conditioned for rotations through π and cannot be inverted at that point. Thus, small discrepancies in λ_{max} are greatly magnified by the inversion of this matrix for angles close to π and from a practical standpoint accuracies in λ_{max} higher than $O(\epsilon^2)$ are needed.

This section first discusses the nature of the exact solutions for the eigenvalues of the K-matrix and shows why the matrix

$$[(\lambda_{max} + \sigma) I - S]$$

becomes ill-conditioned when the angle of rotation passes through π. Following this a number of methods are presented for arriving at a more accurate determination of λ_{max}. These methods are

1. An iterative solution of the characteristic equation for λ_{max}
2. A perturbation expansion for λ_{max}
3. An iterative solution for the Gibbs vector and λ_{max}

Of these, only the first is of practical importance. The second and third methods are given for mathematical completeness and for the insight which they provide on the solutions.

5.1 EXACT SOLUTION OF THE EIGENVALUE PROBLEM

The eigenvalues of K are determined from

$$\det |K - \lambda I| = 0 \tag{5-1}$$

or

$$\det \left[\begin{array}{c|c} (S - (\sigma+\lambda) I) & \vec{z} \\ \hline \vec{z}^{T} & \sigma - \lambda \end{array} \right] = 0 \tag{5-2}$$

Expanding the determinant in minors along the bottom row and right column leads to

$$(\sigma - \lambda) \det |S - (\sigma+\lambda) I| - \sum_{ij} z_i M_{ij} z_j = 0 \tag{5-3}$$

where M_{ij} is the adjoint matrix of $S - (\sigma + \lambda) I$. Recalling the relation between the adjoint and the inverse the above equation may be rewritten as

$$\lambda = \sigma - \vec{z}^{T} \frac{1}{S - (\sigma+\lambda) I} \vec{z} \tag{5-4}$$

Equation (5-4), though computationally identical to Equation (5-1) provides greater insight into the nature of the eigenvalues of K .

Since S is a real symmetric matrix there exists an orthogonal matrix U which induces a similarity transformation diagonalizing S

$$U S U^T = \begin{bmatrix} \Delta_1 & 0 & 0 \\ 0 & \Delta_2 & 0 \\ 0 & 0 & \Delta_3 \end{bmatrix} \tag{5-5}$$

Let

$$\vec{\zeta} = U \vec{z} \tag{5-6}$$

then Equation (5-4) may be rewritten as

$$\lambda = \sigma + \sum_{i=1}^{3} \frac{|\zeta_i|^2}{\lambda + \sigma - \Delta_i} \tag{5-7}$$

Graphically the solution of Equation (5-7) is shown in Figure 5-1.

The right member of Equation (5-7) is a segmented curve with four asymptotes: three verticle lines with abscissas $s_i - \sigma$, $i = 1, 2, 3$, and a horizontal line with ordinate σ . The left member of Equation (5-11) is a straight line of unit slope passing through the origin. The two curves intersect at the points (λ_i , λ_i), $i = 1, 2, 3, 4$.

The following facts are evident from the figure:

1. Equation (5-7) (and therefore also Equation (5-1)) will always have four solutions (including multiple roots)

2. Multiple roots can occur if S has degenerate eigenvalues

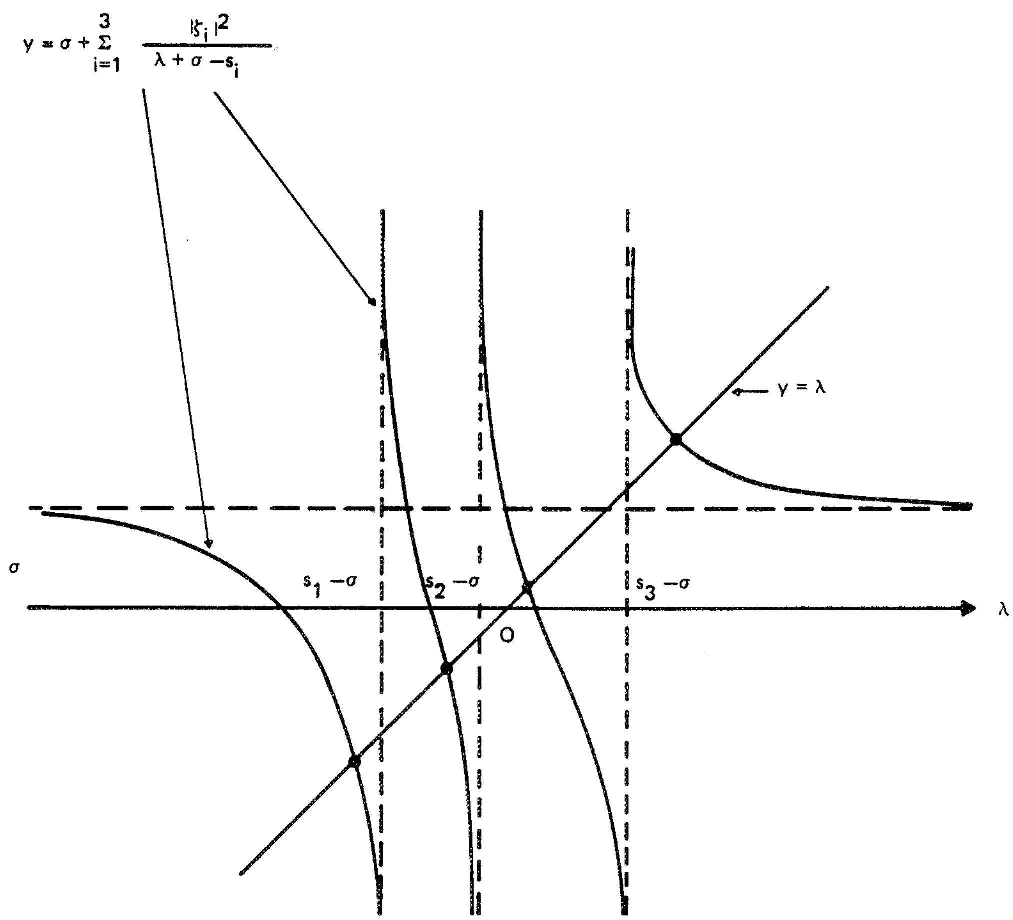


Figure 5-1. Graphical Solution of the Eigenvalue Problem for the Davenport K-Matrix

3. $\lambda_{max} \geq \sigma$

4. Since λ_{max} can never exceed unity the second largest eigenvalue can approach λ_{max} if $s_3 - \sigma$ becomes large enough, i.e., if S has an eigenvalue of magnitude $\lambda_{max} + \sigma$

The question of when S has an eigenvalue $\lambda_{max} + \sigma$ must now be addressed. Recall Equation (4-3), which read

$$\vec{Y} = [(\lambda_{max} + \sigma) I - S]^{-1} \vec{Z} \tag{5-8}$$

If $\hat{X}$ is the axis of the optimal rotation and θ the angle of rotation, then it follows from Equation (2-15) that

$$\vec{Y} = \hat{X} \tan(\theta/2) \tag{5-9}$$

$\vec{Y}$ then is everywhere finite except at $\theta = \pm\pi$. Since the magnitude of $\vec{Z}$ never exceeds unity it follows that one of the eigenvalues of $[(\lambda_{max} + \sigma) I - S]$ must vanish when $\theta = \pm\pi$ and that this matrix is nonsingular for $|\theta| < \pi$.

Thus, for rotations through π it must follow that

$$s_3 = \lambda_{max} + \sigma \tag{5-10}$$

or in terms of the eigenvalues of the K-matrix

Wrong? $s_3 \to 0$

$$\lambda_2 = \lambda_1 = \lambda_{max} \tag{5-11}$$

when the angle of rotation is π.

This does not contradict the result of Section 4.1. Only one eigenvalue of K (the largest eigenvalue, λ_{max}) need be independent of the angle of rotation.

Application of Equations (4-23) and (4-24) to Equation (5-4) leads to a convenient expression (reference 4) for the characteristic equation for λ , namely

$$0 = \det |K - \lambda I| \tag{5-12}$$

or

$$\lambda^4 - (a+b)\lambda^2 - c\lambda + (ab + c\sigma - d) = 0 \tag{5-13}$$

with

$$a = \sigma^2 - \kappa \tag{5-14a}$$

$$b = \sigma^2 + \vec{Z}^T \vec{Z} \tag{5-14b}$$

$$c = \Delta + \vec{Z}^T S \vec{Z} \tag{5-14c}$$

$$d = \vec{Z}^T S^2 \vec{Z} \tag{5-14d}$$

Equations (5-13), (5-14), and (5-8) together with Equations (4-23) and (4-24) provide a very convenient means for computing the optimal rotation (represented by $\vec{Y}$) to arbitrarily high accuracy since the same constants (σ , κ , and Δ) and vectors ($\vec{Z}$, $S\vec{Z}$, and $S^2\vec{Z}$) appear both in the coefficients of the characteristic equation and in the algorithm for inverting the matrix $[(\lambda_{max} + \sigma) I - S]$. λ_{max} can be determined to arbitrarily high accuracy by applying the Newton-Raphson method to Equation (5-13) using as a starting value $\lambda_{max} = 1$.

5.2 PERTURBATIVE SOLUTIONS

Equation (5-4) suggests an obvious approximate expression for λ_{max} when the angle of rotation is small. We introduce the quantity δ defined as

$$\delta = \max(\epsilon, \theta) \tag{5-15}$$

where ϵ is given by Equation (4-7) and θ is the angle of rotation of the optimal rotation. It then follows that

$$\lambda_{max} = \sigma + O(\delta^2) \tag{5-16}$$

Substituting this expression tino the right member of Equation (5-4) leads to

$$\lambda_{max} = \sigma - \vec{Z}^T \frac{1}{S - 2\sigma I} \vec{Z} + O(\delta^4) \tag{5-17}$$

and substituting this expression into Equation (5-8) leads to an expression for $\vec{Y}$ which errs by terms of order δ^5. A more systematic method for obtaining still higher order approximations will now be presented.

Equation (5-15) resembles strongly the result to second order of a perturbation expansion in $\vec{Z}$. In fact, a complete Rayleigh-Schrodinger perturbation expansion (Reference 5) of λ_{max} and $\vec{Y}$ is possible.

Define

$$K_0 = \left[\begin{array}{c|c} S - \sigma I & \vec{0} \\ \hline \vec{0}^T & \sigma \end{array}\right] \tag{5-18a}$$

$$N = \left[\begin{array}{c|c} 0 & \vec{Z} \\ \hline \vec{Z}^T & 0 \end{array}\right] \tag{5-18b}$$

and

$$H_\alpha = K_o + \alpha N \tag{5-19a}$$

Thus,

$$H(0) = K_o \tag{5-19b}$$

$$H(1) = K \tag{5-19c}$$

Note that K_o defined by Equation (5-16a) is different from K_o defined earlier in Section 3.

As a function of α the eigenvalue problem for the quaternion becomes

$$H(\alpha)\,\bar{p}(\alpha) = \lambda(\alpha)\,\bar{p}(\alpha) \tag{5-20}$$

For small rotations $|\vec{Z}|$ is much smaller than the eigenvalues of $S - 2\sigma I$.

It may therefore be expected that an expansion of $\bar{p}(\alpha)$ and $\lambda(\alpha)$

$$\bar{p}(\alpha) = \bar{p}^{(0)} + \alpha\,\bar{p}^{(1)} + \alpha^2\,\bar{p}^{(2)} + \ldots \tag{5-21a}$$

$$\lambda(\alpha) = \lambda^{(0)} + \alpha\,\lambda^{(1)} + \alpha^2\,\lambda^{(2)} + \ldots \tag{5-21b}$$

is possible, which converges in the interval $0 \le \alpha \le 1$. Then

$$\bar{p} = \bar{p}^{(0)} + \bar{p}^{(1)} + \bar{p}^{(2)} + \ldots \tag{5-22a}$$

$$\lambda_{max} = \lambda^{(0)} + \lambda^{(1)} + \lambda^{(2)} + \ldots \tag{5-22b}$$

it being anticipated that for suitable values of $\lambda^{(0)}$ the series will converge to λ_{max} and not to some other eigenvalue of K .

Substituting Equations (5-21ab) into Equation (5-19) and equating terms of equal order in α leads to

$$K_0 \bar{p}^{(0)} = \lambda^{(0)} \bar{p}^{(0)} \tag{5-23a}$$

$$K_0 \bar{p}^{(1)} + N \bar{p}^{(0)} = \lambda^{(0)} \bar{p}^{(1)} + \lambda^{(1)} \bar{p}^{(0)} \tag{5-23b}$$

$$K_0 \bar{p}^{(2)} + N \bar{p}^{(1)} = \lambda^{(0)} \bar{p}^{(2)} + \lambda^{(1)} \bar{p}^{(1)} + \lambda^{(2)} \bar{p}^{(0)} \tag{5-23c}$$

... etc., or

$$K_0 \bar{p}^{(0)} = \lambda^{(0)} \bar{p}^{(0)} \tag{5-24}$$

$$(K_0 - \lambda^{(0)}) \bar{p}^{(k)} + N \bar{p}^{(k-1)} = \sum_{i=1}^{k} \lambda^{(i)} \bar{p}^{(k-i)} \quad , k = 1, 2, \ldots \tag{5-25}$$

From the previous discussion it is obvious that the solution of Equation (5-24) is given by

$$\lambda^{(0)} = \sigma \tag{5-26a}$$

$$\bar{p}^{(0)} = (0, 0, 0, 1)^T \tag{5-26b}$$

Equation (5-25) does not specify the component of $p^{(k)}$ which is parallel to $p^{(0)}$. Therefore, a subsidiary condition may be imposed on $p^{(k)}$

$$\bar{p}^{(0)T} \bar{p}^{(k)} = 0 \tag{5-27}$$

so that $p^{(k)}$ possesses no component along $p^{(0)}$. The operator $(K_o - \lambda^{(0)})$ may then be inverted to give

$$\bar{p}^{(k)} = \frac{1}{K_0 - \sigma I} Q_0 \left\{ -N \bar{p}^{(k-1)} + \sum_{i=1}^{k-1} \lambda^{(i)} \bar{p}^{(k-i)} \right\} \tag{5-28}$$

with

$$Q_0 = I - \bar{p}^{(0)} \bar{p}^{(0)T} \tag{5-29}$$

Operating on Equation (5-25) on the left with $p^{(0)^T}$ yields

$$\lambda^{(k)} = \bar{p}^{(0)T} N \bar{p}^{(k-1)} \tag{5-30}$$

Equations (5-28) through (5-30) are the standard solutions to the Rayleigh-Schrödinger perturbation expansion.

The first few terms of the expansion may be calculated readily. For convenience the following notation will be introduced

$$\frac{1}{K_0 - \sigma I} Q_0 = L = \left[\begin{array}{c|c} (S - 2\sigma I)^{-1} & \vec{0} \\ \hline \vec{0}^T & 0 \end{array} \right] \tag{5-31}$$

Thus,

$$\lambda^{(1)} = \bar{p}^{(0)T} N \bar{p}^{(0)} = 0 \tag{5-32}$$

$$\bar{p}^{(1)} = - N L \bar{p}^{(0)} \tag{5-33}$$

$$\begin{aligned} \lambda^{(2)} &= \bar{p}^{(0)T} N \bar{p}^{(1)} \\ &= - \bar{p}^{(0)T} N L N \bar{p}^{(0)} \\ &= - \vec{Z}^{T} \frac{1}{S - 2\sigma I} \vec{Z} \end{aligned} \tag{5-34}$$

Note that

$$\lambda_{max} = \lambda^{(0)} + \lambda^{(1)} + \lambda^{(2)} + O(\delta^{4}) \tag{5-35}$$

is the same expression as Equation (5-15) above.

Continuing the expansion

$$\bar{p}^{(2)} = - L N \bar{p}^{(1)} = 0 \tag{5-36}$$

$$\lambda^{(3)} = \bar{p}^{(0)T} N \bar{p}^{(2)} = 0 \tag{5-37}$$

$$\bar{p}^{(3)} = L \lambda^{(2)} \bar{p}^{(1)} \tag{5-38}$$

$$\lambda^{(4)} = \bar{p}^{(0)T} N \bar{p}^{(3)} = -\vec{z}^{\,T} \frac{1}{(S-2\sigma I)^2} \vec{z} \cdot \lambda^{(2)} \tag{5-39}$$

$$\begin{aligned} \lambda^{(6)} = & -\vec{z}^{\,T} \frac{1}{(S-2\sigma I)^3} \vec{z} \, [\lambda^{(2)}]^2 \\ & -\vec{z}^{\,T} \frac{1}{(S-2\sigma I)^3} \vec{z} \cdot \lambda^{(4)} \end{aligned} \tag{5-40}$$

$$\begin{aligned} \lambda^{(8)} = & -\vec{z}^{\,T} \frac{1}{(S-2\sigma I)^4} \vec{z} \, [\lambda^{(2)}]^3 \\ & -2\,\vec{z}^{\,T} \frac{1}{(S-2\sigma I)^3} \vec{z} \, \lambda^{(2)} \lambda^{(4)} \\ & -\vec{z}^{\,T} \frac{}{(S-2\sigma I)^2} \vec{z} \, \lambda^{(6)} \end{aligned} \tag{5-41}$$

In general,

$$\bar{p}^{(2m)} = 0 \quad , \quad m = 1, 2, 3, \ldots \tag{5-42}$$

$$\lambda^{(2m+1)} = 0 \quad , \quad m = 0, 1, 2, \ldots \tag{5-43}$$

The kth-order term will contain expressions of the form

$$-\vec{z}^{\,T} \frac{1}{(S-2\sigma I)^m} \vec{z}$$

with $m = 1, 2, \ldots, k$. However, by applying the Cayley-Hamilton theorem of Section 4.2 this expression need be calculated only for $m = 1, 2,$ and 3 , or even more conveniently, for $m = -1, 0,$ and $+1$.

Setting

$$\lambda_{max}^{(2m)} = \lambda^{(0)} + \lambda^{(1)} + \ldots + \lambda^{(2m)} \tag{5-44}$$

with

$$\lambda_{max} = \lambda_{max}^{(2m)} + O(\delta^{2m+2}) \tag{5-45}$$

leads to

$$\vec{Y} = \frac{1}{(\sigma + \lambda_{max}^{(2m)})I - S} \vec{Z} + O(\delta^{2m+3}) \tag{5-46}$$

It may be pointed out that the perturbation series for λ_{max} need not converge necessarily. However, the expansion for λ_{max} can be expected to converge at least asymptotically in the sense that

$$\lambda_{max} = \lambda^{(0)} + \lambda^{(2)} + \ldots + \lambda^{(2m)} + O(\lambda^{(2m+2)}) \tag{5-47}$$

i.e., the error in truncating the perturbation series is of the same order as the first neglected term. For further details on asymptotic series see Reference 5.

The perturbation expansion is of only limited applicability. This is clear from the fact that $\lambda^{(0)} = \sigma$, which in some cases is close to -1 for $\theta = \pi$, while λ_{max} is close to unity. Thus the perturbation expansion will converge very slowly (if at all) for all but very small angles (up to a few degrees). Computation of the perturbation series is, in fact, more complicated and more time consuming than the iterative computation of λ_{max} from the characteristic equation.

5.3 AN ITERATIVE SOLUTION

The perturbation expansion above may be converted to an iterative solution. An iterative solution has the advantage over a perturbation expansion in that convergence properties are usually better and fewer storage locations are required in machine computation. This advantage may be only academic since approximations to Davenport's algorithm are worthwhile only if they require little computation. At some point one is better off to solve the four-dimensional eigenvalue problem explicitly. However, we include the iterative solution for completeness and because it is conceptually very simple.

Define

$$\bar{P}_{(2m+1)} = \bar{p}^{(0)} + \bar{p}^{(1)} + \cdots + \bar{p}^{(2m+1)} \tag{5-48}$$

$$\lambda_{max}^{(2m)} = \lambda^{(0)} + \lambda^{(1)} + \cdots + \lambda^{(2m)} \tag{5-49}$$

Since

$$\bar{p}^{(0)T} \bar{P}_{(2m+1)} = 1 \tag{5-50}$$

for all m it follows that

$$\lim_{m \to \infty} \bar{P}_{(2m+1)} = \lim_{m \to \infty} \begin{Bmatrix} \vec{P}_{2m+1} \\ 1 \end{Bmatrix} = \begin{Bmatrix} \vec{Y} \\ 1 \end{Bmatrix} \tag{5-51}$$

at least in the asymptotic sense defined in Section 5.2. Clearly,

$$\vec{Y} = \vec{P}_{(2m+1)} + O(\delta^{2m+3}) \tag{5-52}$$

and

$$\lambda_{max}^{(2m+2)} = \sigma + \vec{Z} \cdot \vec{P}_{(2m+1)} + O(\delta^{2m+4}) \tag{5-53}$$

Define

$$\vec{Y}^{(2m+1)} = \left\{(\sigma + \lambda_{max}^{(2m)})I - S\right\}^{-1} \vec{Z} \tag{5-54}$$

whence from Equation (5-46)

$$\vec{P}_{(2m+1)} = \vec{Y}^{(2m+1)} + O(\delta^{2m+3}) \tag{5-55}$$

Combining Equations (5-52) through (5-55) leads to the following set of recursion relations

$$\lambda_{max}^{(0)} = \sigma \tag{5-56}$$

$$\vec{Y}^{(2m+1)} = \left\{(\sigma + \lambda_{max}^{(2m)})I - S\right\}^{-1} \vec{Z} \tag{5-57}$$

$$\lambda_{max}^{(2m+2)} = \sigma + \vec{Z} \cdot \vec{Y}^{(2m+1)} \tag{5-58}$$

While the perturbation expansion of Section 5.2 may be unstable, Equations (5-56) through (5-58) will have stable solutions at least for small angles of rotation. This can be seen from the fact that an increase in $\lambda_{max}^{(2m)}$ with respect to $\lambda_{max}^{(2m-2)}$ will decrease $\vec{Y}^{(2m+1)}$ relative to $\vec{Y}^{(2m-1)}$. Hence $\lambda_{max}^{(2m+2)}$ will be smaller than $\lambda_{max}^{(2m)}$.

The computation of Equations (5-57) and (5-58) can be simplified by using the Cayley-Hamilton theorem as in Section 4.2. Define

$$\alpha_{2m} = \alpha(\lambda_{max}^{(2m)}) \tag{5-59a}$$

$$\beta_{2m} = \beta(\lambda_{max}^{(2m)}) \tag{5-59b}$$

$$\gamma_{2m} = \gamma(\lambda_{max}^{(2m)}) \tag{5-59c}$$

Then

$$\vec{Y}^{(2m+1)} = \gamma_{2m}^{-1}\left[\alpha_{2m}\vec{Z} + \beta_{2m} S\vec{Z} + S^2\vec{Z}\right] \tag{5-60}$$

$$\lambda_{max}^{(2m+2)} = \sigma + \gamma_{2m}^{-1}\left[\alpha_{2m}\vec{Z}^T\vec{Z} + \beta_{2m}\vec{Z}^T S\vec{Z} + \vec{Z}^T S^2\vec{Z}\right] \tag{5-61}$$

which has the advantage that the matrix operations need be performed only once. Equations (5-59) and (5-61) define a recursive algorithm whereby the error decreases by two orders of magnitude (from $O(\delta^{2m})$ to $O(\delta^{2m+2})$) with each cycle. It is therefore less efficient than applying the Newton-Raphson method to find the root near 1 of the characteristic polynomial given by Equation (5-13) and also demands considerably more computation.

There is no reason to use $\lambda_{max}^{(0)} = \sigma$ as the initial value of λ_{max} in the iterative loop of Equations (5-56), (5-57), and (5-58) except to maintain an artificial and unnecessary resemblance to the rather cumbrous perturbation series. The iterative method will have wider application and converge more quickly if instead the choice

$$\lambda_{max}^{(0)} = 1 \tag{5-56'}$$

is made.

It should be noted that Equation (5-58) is not well adapted to computing λ_{max} for rotation angles near $\theta = \pi$ since as $\theta \to \pi$

$$|\vec{Z}| \to 0 \tag{5-62a}$$

$$|\vec{Y}| \to \infty \tag{5-62b}$$

but

$$\vec{Z} \cdot \vec{Y} \to 1 - \sigma + O(\varepsilon^2) \tag{5-62c}$$

SECTION 6 - NUMERICAL RESULTS AND DISCUSSION

The preceding text offered a large number of algorithms by which optimal solutions for spacecraft attitude can be generated. The problem now is which one to choose in actual application.

One may immediately discard for most applications those algorithms which are applicable to only small angles such as Equations (3-27), (3-28), (3-47), (4-24), (4-25), and the perturbation expansion of Section 5 since these are for the most part no simpler to implement than replacing λ_{max} with unity in the exact formula.

An exception may be made in two cases. In the first case one may have some foreknowledge of the spacecraft attitude. The difference between this a priori attitude and the true attitude may be sufficiently small that these small-angle algorithms are useful. The computational advantage, however, seems very slight compared to calculating the true attitude directly by a more exact method without using an intermediate attitude solution unless the analytic structure of these intermediate attitude solutions is particularly simple.

In the other case one may imagine the spacecraft to have some high pointing accuracy in inertial space. In that case the angle of rotation is always small and the matrix S is always very close to the matrix S_o of Section 3 for all allowed orientations of the spacecraft. The matrix $[2I - S_o]^{-1}$ need then only be computed once for all times and the vector components of the quaternion are obtained immediately by the same linear transformation of the vector $\vec{Z}$. The amount of core storage required under these circumstances would be very small and the algorithm of Equation (3-28) could easily be implemented onboard.

The Magsat spacecraft rotates about its pitch axis at a rate of about 4 arc minutes per second and, thus, one is led to examine the more exact and universally applicable algorithms. This limits the choice to using the Equation (4-22) with one of three possible choices for λ_{max}. These are

1. $\lambda_{max} = 1$
2. λ_{max} is obtained by applying the Newton-Raphson method to the characteristic Equation (5-3) using $\lambda_{max} = 1$ as a starting value
3. λ_{max} is obtained by the iterative method of Equations (5-59) and (5-61) using $\lambda_{max} = 1$ as a starting value.

The third choice may be eliminated since it is more cumbersome to implement than the second and not more accurate. Therefore, only the first and second choices have been examined numerically for a large number of possible rotations.

The first two choices above are, in fact, the same method. The first choice is simply the zeroth iteration of the Newton-Raphson method applied to Equation (5-3), with a starting value of $\lambda_{max} = 1$. It will be more convenient, therefore, not to speak of choice 1 or 2 but rather of the order of iteration of the Newton-Raphson method.

These attitude computation algorithms have been tested with a sensor configuration approximating that of the Magsat spacecraft. This includes the following sensors and orientations:

Sensor	Coelevation Degrees	Azimuth Degrees
Fine Sun sensor	180	-
Star camera	60	45
Star camera	60	135

The coordinate system in which these sensor orientations are defined is given in the Magsat Fine Attitude Determination Study (Reference 6). The results of this section, of course, are independent of the choice of coordinate system. In this coordinate system the three model reference unit-vectors are

$$\hat{V}_1 = (0, 0, 1)^T \tag{6-1a}$$

$$\hat{V}_2 = (\sqrt{3/8}, \sqrt{3/8}, 1/2)^T \tag{6-1b}$$

$$\hat{V}_3 = (-\sqrt{3/8}, \sqrt{3/8}, 1/2)^T \tag{6-1c}$$

The smallest angle between any two vectors is 60 degrees; thus, the condition that the reference vectors (or the observation vectors) not be all collinear is well satisfied.

The algorithm for computing the optimal attitudes were subjected to two tests, first for input data where the $\hat{W}_i$ were chosen such that

$$\hat{W}_i = R(\bar{\xi}_{true})\hat{V}_i \ , \ i = 1, 2, 3 \tag{6-2}$$

and secondly for input data where the data were chosen such that

$$\hat{W}_i = R(\bar{\xi}_{true})\hat{U}_i \ , \ i = 1, 2, 3 \tag{6-3}$$

$$\hat{U}_i = R_0^{-1}(\hat{V}_i + \delta\vec{V}_i)/|\hat{V}_i + \delta\vec{V}_i| \ , \ i = 1, 2, 3 \tag{6-4}$$

The $\delta\vec{V}_i$ were chosen arbitrarily to have the values

$$\delta\vec{V}_1 = (\varepsilon, 0, 0)^T \tag{6-5a}$$

$$\delta\vec{V}_2 = (0, \varepsilon, 0)^T \tag{6-5b}$$

$$\delta\vec{V}_3 = (0, 0, \varepsilon)^T \tag{6-5c}$$

with

$$\varepsilon = 5 \times 10^{-5} \tag{6-6}$$

so that compared to $\hat{V}_i$ each $\hat{U}_i$ has an error equivalent to a rotation of about 10 arc-seconds. The matrix R_o^{-1} is chosen so that the optimal rotation which carries the $\hat{U}_i$ into the $\hat{V}_i$ is the identity matrix. If this specific choice were not made, it would be difficult to define precisely what is meant by a rotation through a given angle.

In each test a true quaternion $\bar{q}_{true}$ was constructed for a given θ and $\hat{X}$ according to Equation (2-15) and the matrix $R(\bar{q}_{true})$ was constructed according to Equation (2-17). Then, given the $\hat{W}_i$ and $\hat{V}_i$, the algorithm was used to compute $\bar{q}_{test}$. The accuracy of the algorithm was determined by computing

$$D = 2\,|\vec{Q}_{true} - \vec{Q}_{test}| \tag{6-7}$$

which ought to be much smaller than ϵ.* In terms of the angle of rotation, θ, D has the approximate value

$$D = \left|\sin\left(\theta_{true}/2\right)\right| \cdot |\Delta\theta| \tag{6-8}$$

where $\Delta\theta$ is the angle of rotation from the true attitude solution to the test attitude solution. Thus, for true angles of rotation near π, where the algorithm meets the most difficult test,

$$D \approx |\Delta\theta| \tag{6-9}$$

For the first test (for which $\lambda_{max} = 1$ exactly) it was found for many different choices of the true axis of rotation $\hat{X}$ that the algorithm gave solutions which were accurate to within 1 arc-second so long as

$$|\pi - \theta_{true}| \gtrsim 10^{-7} \approx .02 \text{ arc seconds}^{**} \tag{6-10}$$

For values of $|\pi - \theta_{true}|$ smaller than this value, the accuracy of the results is much poorer and there is the possibility of encountering overflows. Overflows occur at these values of θ_{true} because a double precision real variable in FORTRAN has only 16.8 significant decimal places. The cause of the overflow is the division by γ*** (see Equation (4-22)) which tends to zero as $\theta_{true} \rightarrow \pi$. Since γ depends on the difference of quantities which differ by $O\left(|\pi - \theta_{true}|^2\right)$, it becomes very probable that the value of γ will be truncated

---

*$\bar{q}_{true}$ and $\bar{q}_{test}$ were always chosen to have the fourth component non-negative.

**When no units are given for angular measure these may be assumed to be radians.

***In Section 6, γ always means $\gamma(\lambda_{max})$, likewise for α and β.

to 0 when $|\pi - \theta_{true}| \lesssim 10^{-8}$. If single precision variables were used, these overflows would be expected to occur when $|\pi - \theta_{true}| \lesssim 10$ arc-seconds, which for Magsat is not acceptable.

The second test with "noised" observation vectors requires some detailed discussion. For $\lambda_{max} = 1$ (no iterations) typical errors were 10^{-10} and the algorithm began to break down for $|\pi - \theta| < 10^{-5}$ at which point the computational error was already as large as 1 arc-second.

For one iteration of λ_{max} the quality of the algorithm improved dramatically. Typical computation errors were 10^{-15} radians for angles no greater than 179.5 degrees. For $|\pi - \theta_{true}| \approx 10^{-8}$ the error was no worse than 10^{-8} (≈ 0.002 arc-second), after which the algorithm began to break down or even overflow.

For two or more iterations of λ_{max} the results were not quite as good. Typical computation errors were 10^{-10} and became as large as 0.3 arc-second for $|\pi - \theta_{true}| = 10^{-6}$ (≈ 0.2 arc-second), after which they became much worse.

The greater success of the algorithm with only one iteration of λ_{max} can be understood in terms of the size of ϵ and the truncation error of the computer. The least significant bit for a double precision word in FORTRAN is equivalent to $10^{-16.8}$ in decimal notation while $\epsilon \approx 10^{-4.3}$. Thus, $10^{-16.8} = \epsilon^{3.9}$, which is very close to ϵ^4, the order of the "algebraic" error in λ_{max} after one iteration of the Newton-Raphson method. The order of the "algebraic" error after the next iteration is ϵ^8, which is lost in the truncation error of the computer. Thus, the accuracy of computing λ_{max} cannot be improved by further iteration of the Newton-Raphson method. If there is a close neighboring root to the one sought, further iteration may even degrade the result. This is indeed what happens when θ_{true} approaches π.

The optimum order of iteration is a function of the computer and the "noise" level of the input observation vectors. For $\epsilon = 10^{-2}$ (≈ 0.5 degree) two iterations would probably give the best results. Had single precision been chosen for testing the algorithm for $\epsilon = 5 \times 10^{-5}$ above, the value $\lambda_{max} = 1$, i.e., zero iterations, would have yielded results which could not have been improved by further iteration. In each case the optimal order should be obtained from trial computations like the ones performed above.

The probability of overflows at angles near π may be decreased by computing $\bar{q}$ somewhat differently. This also improves the accuracy of the solutions near $\theta = \pi$.

Note that

$$\vec{Y} = \gamma^{-1} \vec{X} \qquad (6\text{-}11)$$

with

$$\vec{X} = \alpha \vec{z} + \beta S \vec{z} + S^2 \vec{z} \qquad (6\text{-}12)$$

In general, $|\vec{X}|$ is O(1) for angles of rotation close to π. Exceptions to this can occur as noted below.

From Equation (6-11) it follows that

$$\vec{Q} = \vec{X} / \sqrt{\gamma^2 + |\vec{X}|^2} \qquad (6\text{-}13a)$$

$$\xi = \gamma / \sqrt{\gamma^2 + |\vec{X}|^2} \qquad (6\text{-}13b)$$

Using Equation (6-13) rather than Equations (6-11) and (4-4) restrict the possibility of overflow to those cases where γ and all three components of $\vec{X}$ vanish simultaneously.

It was shown in Section 5.1 that the matrix $[(\lambda_{max} + \sigma) I - S]$ had a vanishing eigenvalue only for rotations through π. From

$$\gamma = \det |(\lambda_{max} + \sigma) I - S| \tag{6-14}$$

it follows that γ vanishes only $\theta = \pi$.

From Equation (6-11)

$$|\gamma| = |\vec{X}| \cdot \left|\tan\left(\frac{\pi - \theta}{2}\right)\right| \tag{6-15}$$

Since $0 < |\vec{Y}| < \infty$ for $0 < \theta < \pi$ it follows that $\vec{X}$ can vanish only for $\theta = 0$ or $\theta = \pi$. $\vec{X}$ must vanish for $\theta = 0$ since $\vec{Q}$ vanishes there. $\vec{X}$ can also vanish at $\theta = \pi$ only if the rotation satisfies certain subsidiary conditions, which we now determine.

Note first that $\vec{X}$ can vanish if and only if $\vec{Z}$ vanishes. This follows from

$$\vec{X} = (\alpha I + \beta S + S^2)\,\vec{Z} \tag{6-16}$$

and

$$\alpha I + \beta S + S^2 = \frac{\det |(\lambda_{max} + \sigma) I - S|}{(\lambda_{max} + \sigma) I - S} \tag{6-17}$$

The matrix $[\alpha I + \beta S + S^2]$ is clearly unimodular and, therefore, nonsingular. Hence, $\vec{X}$ vanishes if and only if $\vec{Z}$ vanishes and it is sufficient to determine the requirement for the vanishing of $\vec{Z}$ for rotations through π.

Neglecting the errors of observation

$$\vec{Z} = \sum_i a_i \, (R\hat{V}_i) \times \hat{V}_i \tag{6-18}$$

If $\hat{n}$ is the axis of rotation then $\hat{V}_i$ may be decomposed as

$$\hat{V}_i = \vec{V}_i^{\parallel} + \vec{V}_i^{\perp} \tag{6-19}$$

with

$$\vec{V}_i^{\parallel} = \hat{n}(\hat{n} \cdot \hat{V}_i) \tag{6-20a}$$

$$\vec{V}_i^{\perp} = -\hat{n} \times (\hat{n} \times \hat{V}_i) \tag{6-20b}$$

For a rotation through π

$$R\hat{V}_i = \vec{V}_i^{\parallel} - \vec{V}_i^{\perp} \tag{6-21}$$

whence

$$\vec{Z} = 2 \sum_i a_i \, \vec{V}_i^{\parallel} \times \vec{V}_i^{\perp}$$

$$= 2 \sum_i a_i (\hat{V}_i \cdot \hat{n})(\hat{n} \times \hat{V}_i) \tag{6-22}$$

or

$$\vec{Z} = \hat{n} \times (S_0 \hat{n}) \tag{6-23}$$

with

$$S_0 = 2 \sum_i a_i \hat{V}_i \hat{V}_i^T \tag{6-24}$$

the matrix which was introduced in Section 3.2.

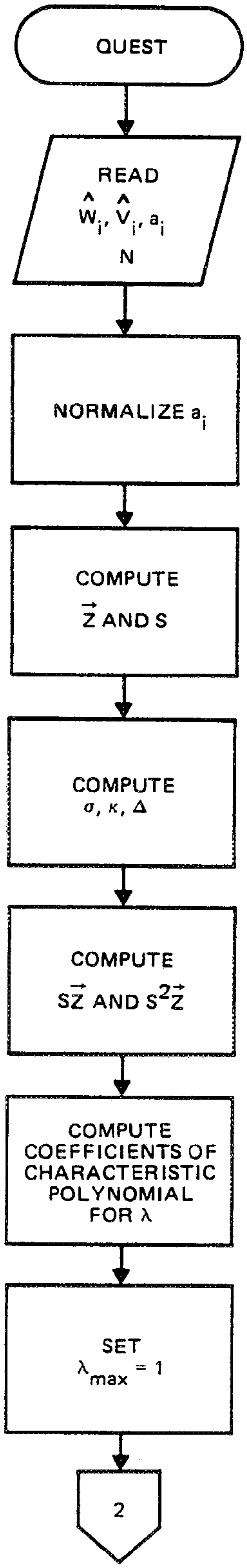
Thus, $\vec{Z}$, and hence $\vec{X}$, vanishes for a rotation through π if and only if the axis of rotation $\hat{n}$ is an eigenvector of S_o . If S_o is proportional to the identity matrix (which occurs in the case where there are three mutually orthogonal reference vectors with equal weights) this condition will always be satisfied. For the case of Magsat the eigenvalues of S_o are nondegenerate and the probability of $\hat{n}$ being an eigenvector of S_o is much reduced.

It might be pointed out that even in those cases where no overflow occurs at $\theta = \pi$ the accuracy of the method will still be poor very near $\theta = \pi$ since much information is lost in the cancellation of large nearly equal numbers.

This algorithm has been tested extensively for the Magsat sensor configuration with λ_{max} calculated from one iteration of the Newton-Raphson method. Overflows were almost eliminated using the methods just discussed. It was found for all choices of the true axis of rotation that the computation error was on the order of 10^{-15} radians ($\approx 2 \times 10^{-10}$ arc-seconds) for $0 \lesssim \theta_{true} \lesssim 179.5$ degrees . Thereafter, for each decade decrease in $|\pi - \theta_{true}|$ the computational error of the solution increased by one decade. This is the expected behavior for a purely truncational error. This law was found to hold until the value $|\pi - \theta_{true}| \approx 10^{-15}$, beyond which $|\pi - \theta_{true}|$ is essentially truncated to zero. The error becomes greater than 1 arc-second for $|\pi - \theta_{true}| \lessapprox 10^{-11}$. Taking 2 arc-seconds as the limit of acceptable computational error for the Magsat mission (attitude accuracy requirement = 20 arc-seconds) and noting that the attitude will be computed four times per second during the Magsat mission, it can then be expected that one unacceptable attitude computation will be encountered every 50,000 years. This is significantly larger than the estimated Magsat mission lifetime of from four to eight months. During the lifetime of the Magsat mission it is, in fact, expected that no more than one attitude computation will have a computational error larger than 10^{-4} arc-seconds. Overflows due to the vanishing of $\gamma^2 + |\vec{X}|^2$ should occur no more than once in 10^{20} years. No alternative computational methods (such as those described below) are planned for Magsat.

The logical flow of the calculation of the optimal quaternion is given in Figure 6-1. The subroutine which performs these operations is called QUEST (for "quaternion estimator").

Cases may nonetheless arise in other missions when it is necessary to avoid the singularity at $\theta = \pi$. This might happen if it were necessary to work in single precision or implement the algorithm for an onboard processor whose words have even fewer bits. In this case it may be noted that if the angle of rotation is greater than $\pi/2$, then the optimal rotation can be expressed as a rotation through π about one of the coordinate axes followed by a rotation through an angle less than $\pi/2$. A rotation through π about one of the coordinate axes will only change the sign of two of the components of each reference vector. The quaternion $\bar{p} = (p_1, p_2, p_3, p_4)^T$ of the rotation transforming the new reference vectors $\hat{V}'_i$ to the observation vectors $\hat{W}_i$ is related very



$$\sum_i a_i = 1$$

$$\vec{Z} = \sum_i a_i \hat{W}_i \times \hat{V}_i$$

$$S = \sum_i a_i \left[\hat{W}_i \hat{V}_i^T + \hat{V}_i \hat{W}_i^T \right]$$

$$\sigma = \tfrac{1}{2} \operatorname{Tr} S$$

$$\kappa = \operatorname{Tr} (\operatorname{adj} S)$$

$$\Delta = \det S$$

$$f(\lambda) = \lambda^4 - (a + b)\lambda^2 - c\lambda + (ab + c\sigma - d) = 0$$

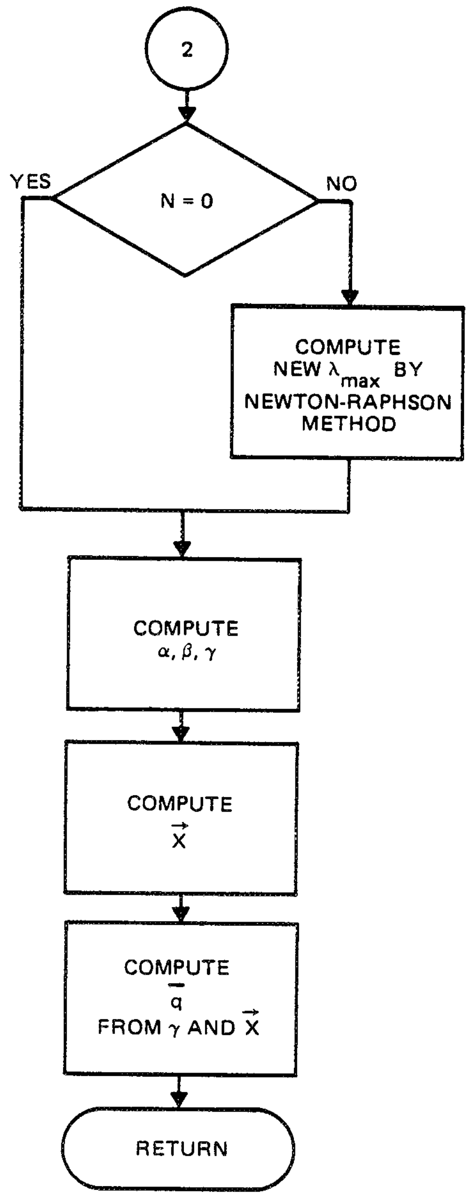
$$a = \sigma^2 - \kappa$$

$$b = \sigma^2 + \vec{Z}^T \vec{Z}$$

$$c = \Delta + \vec{Z}^T S \vec{Z}$$

$$d = \vec{Z}^T S^2 \vec{Z}$$

Figure 6-1. Logical Flow for the Computation of the Optimal Quaternion (1 of 2)



$$\lambda'_{max} = \lambda_{max} + \delta\lambda_{max}$$

$$\delta\lambda_{max} = -\frac{f(\lambda_{max})}{f'(\lambda_{max})}$$

ITERATE N TIMES

$$\alpha = \lambda_{max}^2 - \sigma^2 + \kappa$$

$$\beta = \lambda_{max} - \sigma$$

$$\gamma = (\lambda_{max} + \sigma)(\lambda_{max}^2 - \sigma^2 + \kappa) - \Delta$$

$$\vec{X} = \alpha \vec{Z} + \beta S \vec{Z} + S^2 \vec{Z}$$

$$\vec{Q} = \vec{X} / \sqrt{\gamma^2 + |\vec{X}|^2}$$

$$q = \gamma / \sqrt{\gamma^2 + |\vec{X}|^2}$$

Figure 6-1. Logical Flow for the Computation of the Optimal Quaternion (2 of 2) simply to the desired quaternion $\bar{q} = (q_1, q_2, q_3, q_4)^T$ for the total rotation. The results are as follows

1. 180-degree rotation about x-axis

   $\hat{V}'_i = (\hat{V}_{ix}, -\hat{V}_{iy}, -\hat{V}_{iz})^T$

   $q_1 = p_4 , \; q_2 = -p_3 , \; q_3 = p_2 , \; q_4 = -p_1$

2. 180-degree rotation about y-axis

   $\hat{V}'_i = (-\hat{V}_{ix}, \hat{V}_{iy}, -\hat{V}_{iz})$

   $q_1 = p_3 , \; q_2 = p_4 , \; q_3 = -p_1 , \; q_4 = -p_2$

3. 180-degree rotation about z-axis

   $\hat{V}'_i = (-\hat{V}_{ix}, -\hat{V}_{iy}, \hat{V}_{iz})$

   $q_1 = -p_2 , \; q_2 = p_1 , \; q_3 = p_4 , \; q_4 = -p_3$

It might be pointed out as a final remark that λ_{max} has the value

$$\lambda_{max} = 1 - \frac{1}{2} \sum_{i=1}^{n} a_i \| \hat{W}_i - R_{opt} V_i \|^2 \tag{6-25}$$

where R_{opt} is the optimal rotation. Thus, if the "noise" in the observation (or reference) vectors is known to have an amplitude ϵ , any significant deviation of $1 - \lambda_{max}$ from ϵ^2 would indicate poor input data, for example, the misidentification of a star.

SECTION 7 - SUMMARY

This section summarizes the analysis for the attitude computation algorithm proposed for the Magsat mission fine attitude determination system.

A fast and accurate method has been developed for computing an optimal rotation, R_{opt} , which carries a set of n reference vectors $V_1, \ldots, V_n$, into a set of n observation vectors $W_1, \ldots, W_n$. This optimal rotation minimizes a weighted square loss function

$$\ell(R) = \frac{1}{2} \sum_{i=1}^{n} a_i \left| \hat{W}_i - R \hat{V}_i \right|^2 \tag{7-1}$$

where

$$a_i \geq 0 \qquad i = 1, \ldots, n \tag{7-2}$$

and

$$\sum_{i=1}^{n} a_i = 1 \tag{7-3}$$

It is assumed also that at least two of the a_i are nonvanishing and the corresponding reference vectors are not parallel (or antiparallel). If the $\hat{W}_i$, i = 1, ..., n , are the representations of the $\hat{V}_i$, i = 1, ..., n , in the spacecraft body system, then R_{opt} is the optimal estimate of the spacecraft attitude in the reference frame of the $\hat{V}_i$. The method presented here is based on the q-method of Davenport (References 1 and 2).

The rotation matrix R is related to the quaternion representation of the rotation, $\bar{q} = (q_1, q_2, q_3, q_4)^T$, according to

$$R(\bar{q}) = \begin{bmatrix} q_1^2 - q_2^2 - q_3^2 + q_4^2 & 2(q_1 q_2 + q_3 q_4) & 2(q_1 q_3 - q_2 q_4) \\ 2(q_1 q_2 - q_3 q_4) & -q_1^2 + q_2^2 - q_3^2 + q_4^2 & 2(q_2 q_3 + q_1 q_4) \\ 2(q_1 q_3 + q_2 q_4) & 2(q_2 q_3 - q_1 q_4) & -q_1^2 - q_2^2 + q_3^2 + q_4^2 \end{bmatrix} \tag{7-4}$$

In terms of $\bar{q}$, the loss function is

$$l(\bar{q}) \equiv l(R(\bar{q})) = 1 - g(\bar{q}) \tag{7-5}$$

with

$$g(\bar{q}) = \bar{q}^T K \bar{q} \tag{7-6}$$

The 4 x 4 matrix K has the form

$$K = \left[\begin{array}{c|c} S - \sigma I & \vec{Z} \\ \hline \vec{Z}^T & \sigma \end{array}\right] \tag{7-7}$$

where

$$S = \sum_{i=1}^{n} a_i \left(\hat{W}_i \hat{V}_i^T + \hat{V}_i \hat{W}_i^T\right) \tag{7-8}$$

$$\sigma = \frac{1}{2} \operatorname{Tr} S \tag{7-9}$$

$$\vec{Z} = \sum_{i=1}^{n} a_i \left(\hat{W}_i \times \hat{V}_i \right) \tag{7-10}$$

The optimal quaternion which minimizes $\ell(\bar{q})$ is given by

$$\bar{q} = \begin{Bmatrix} \vec{Q} \\ q \end{Bmatrix} = \frac{1}{\sqrt{1+|\vec{Y}|^2}} \begin{Bmatrix} \vec{Y} \\ 1 \end{Bmatrix} \tag{7-11}$$

with

$$\vec{Y} = \left\{ (\lambda_{max} + \sigma) I - S \right\}^{-1} \vec{Z} \tag{7-12}$$

where λ_{max} is the largest eigenvalue of K.

$$\vec{Y} = \vec{Q}/q \tag{7-13}$$

is the Gibbs vector of the rotation.

If an exact simultaneous rotation of the $\hat{V}_i$ into the $\hat{W}_i$ is possible, then λ_{max} has the value unity. Otherwise λ_{max} is smaller than unity by the weighted sum square of the residuals

$$\lambda_{max} = 1 - \frac{1}{2} \sum_{i=1}^{n} a_i \left| \hat{W}_i - R_{opt} \hat{V}_i \right|^2 \tag{7-14}$$

The magnitude of the residuals $(\hat{W}_i - R_{opt} \hat{V}_i)$ will be of the order of the error (in radian measure) of the spacecraft sensors determining the observation vector. (The error in the reference vectors is assumed to be much smaller).

Thus, λ_{max} is smaller than unity by half the mean sum square of the observation errors (in radians).

If the value

$$\lambda_{max} \approx 1 \tag{7-15}$$

is substituted in Equation (7-12) an estimate is obtained for the quaternion which is in error by an amount of the same order as the mean sum square of the sensor errors.

λ_{max}, $\vec{Y}$ and $\bar{q}$ can be determined to arbitrarily high accuracy as follows: one defines the quantities

$$\begin{aligned} \kappa &= Tr\,(adj\; S) \\ &= S_{11}S_{22} + S_{22}S_{33} + S_{33}S_{11} \\ &\quad - S_{12}S_{21} - S_{23}S_{32} - S_{31}S_{13} \end{aligned} \tag{7-16}$$

$$\Delta = \det S \tag{7-17}$$

$$\alpha = \lambda_{max}^{2} - \sigma^{2} + \kappa \tag{7-18}$$

$$\beta = \lambda_{max} - \sigma \tag{7-19}$$

$$\gamma = (\lambda_{max} + \sigma)\,\alpha - \Delta \tag{7-20}$$

Then λ_{max} is a root of

$$\lambda^4 - (a+b)\lambda^2 - c\lambda + (ab + c\sigma - d) = 0 \tag{7-21}$$

with

$$a = \sigma^2 - \varkappa \tag{7-22}$$

$$b = \sigma^2 + \vec{Z}^T \vec{Z} \tag{7-23}$$

$$c = \Delta + \vec{Z}^T S \vec{Z} \tag{7-24}$$

$$d = \vec{Z}^T S^2 \vec{Z} \tag{7-25}$$

which can be solved to arbitrarily high accuracy using the Newton-Raphson method and a starting value $\lambda_{max} = 1$. For observation errors less than one degree the limiting accuracy of a double-precision word in IBM FORTRAN will be attained after two iterations. One iteration will suffice for observation errors less than 10 arc-seconds.

In terms of the above, the Gibbs vector may be written as

$$\vec{Y} = \gamma^{-1} \vec{X} \tag{7-26}$$

with

$$\vec{X} = \alpha \vec{Z} + \beta S \vec{Z} + S^2 \vec{Z} \tag{7-27}$$

The quaternion may likewise be written as

$$\bar{q} = \frac{1}{\sqrt{\gamma^2 + |\vec{X}|^2}} \begin{Bmatrix} \vec{X} \\ \gamma \end{Bmatrix} \tag{7-28}$$

This last formula is more accurate than obtaining $\bar{q}$ using Equation (7-11) as an intermediate step.

The quantity $(\gamma^2 + |\vec{X}|^2)$ will vanish when the angle of rotation is π and the axis of rotation is an eigenvector of S_o with

$$S_o = 2 \sum_{i=1}^{n} a_i \hat{V}_i \hat{V}_i^T \tag{7-29}$$

A procedure was presented at the end of Section 6, whereby this situation can always be avoided. When this is done the accuracy of the method is very nearly the accuracy with which real constants can be represented in the computer. For Magsat the likelihood of encountering overflows or even large computational errors is sufficiently small that these additional procedures need not be implemented.

REFERENCES

1. P. Davenport (private communication)

2. Computer Sciences Corporation, CSC/TM-77/6034, Analysis of Least Squares Attitude Determination Routine, DOAOP, J. Keat, February 1977

3. F. L. Markley, "Attitude Dynamics," Chapter 16 in Attitude Determination and Control, edited by J. R. Wertz. Doordrecht, The Netherlands: D. Reidel Publishing Company (in press)

4. F. L. Markley (private communication)

5. A. Nayfeh, Perturbation Methods, New York, John Wiley & Sons (1973)

6. Computer Sciences Corporation, CSC/TM-77/6203, Magsat Fine Attitude Determination Study, D. M. Gottlieb and M. D. Shuster, October 1977

7. G. M. Lerner, "Deterministic Attitude Methods," Chapter 12 in Attitude Determination and Control, edited by J. R. Wertz. Doordrecht, The Netherlands: D. Reidel Publishing Company (in press)